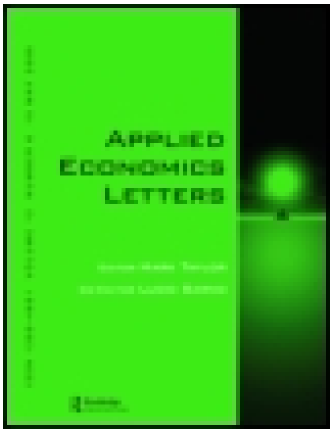

Applied Economics Letters
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rael20

A new method to choose optimal lag order in stable and unstable VAR models

A. Hatemi-J [a]

[a] Department of Economics, Lund University and Department of Economics and Political Sciences, University of Skövde, PO Box 408, SE-541 28, Skövde, Sweden. E-mail: abdulnasser.hatemi-j@ish.his.se

Published online: 06 Oct 2010.

Applied Economics Letters, 2003, **10**, 135–137

A new method to choose optimal lag order in stable and unstable VAR models

A. HATEMI-J

Department of Economics, Lund University and Department of Economics and Political Sciences, University of Skövde, PO Box 408, SE-541 28, Skövde, Sweden
E-mail: abdulnasser.hatemi-j@ish.his.se

A crucial aspect of empirical research based on the vector autoregressive (VAR) model is the choice of the lag order, since all inference in the VAR model is based on the chosen lag order. Here, a new information criterion is introduced for this purpose. The conducted Monte Carlo simulation experiments show that this new information criterion performs well in picking the true lag order in stable as well as unstable VAR models.

I. INTRODUCTION

The vector autoregressive (VAR) model has been one of the most applied models in the field of time series analysis.[1] This model allows for interaction between the variables of interest. With new advancements in the field of time series econometrics, especially when the data generating process (DGP) is characterized by unit roots, the VAR model (or its vector error correction representation) has become even more of use because it allows for exploring the long-run relations between the variables in combination with short-run potential dynamics. The VAR model is also intensively used to test for causality in the Ganger's sense. It is also possible to investigate the effect of policy changes by using moving average representation of the VAR models to calculate impulse response functions and variance decompositions. The VAR model is also known to have good forecasting properties.

A crucial aspect in this regard is the choice of the optimal lag order because all inference in the VAR model is based on the chosen lag order. There are several information criteria in the literature that offer the possibility of choosing the optimal lag order.[2] According to simulation results, Schwarz (1978) Bayesian information criterion and the Hannan and Quinn (1979) information criterion seem to perform best.[3] Sometimes these criteria can choose different lag orders. The question is then upon which of these criteria should one rely. This article suggests combining these two criteria to obtain a new criterion for choosing the lag order. The conducted simulation experiments show that this simple criterion is successful in picking the true lag order in both stable and unstable VAR models.[4]

This article is organized as follows. Section II presents the VAR model and a new information criterion for picking the right lag order. Section III introduces the design of the simulation procedure. Section IV provides the simulation results and the last section concludes.

II. THE OPTIMAL LAG ORDER IN THE VAR MODEL

Let us concentrate on a multivariate time series, X, consisting of n variables that is characterized by a VAR model of an order less-than or equal to K:

$$X_t = \Gamma D_t + \sum_{k=1}^{K} \beta_k X_{t-k} + \varepsilon_t, \qquad t = 1, \ldots, T \tag{1}$$

[1] The VAR model was introduced by Sims (1980).
[2] The first information criterion for this purpose was introduced by Akaike.
[3] See Lütkepohl (1985) and Hacker and Hatemi-J (2001).
[4] The Monte Carlo simulations were performed using Gauss.

Applied Economics Letters ISSN 1350–4851 print/ISSN 1466–4291 online © 2003 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/1350485022000041050

where ε_t is a $n \times 1$ vector of disturbances that are assumed to be independently identically distributed errors with the distribution $N_n(0, \Omega)$, β_k is a matrix of coefficients for X_{t-k}, and D_t signifies non-stochastic components such as constant terms, linear trend, or seasonal dummies. The initial values, $X_{1-K}, \ldots, X_0$, are assumed to be fixed. The objective is to choose the largest order for the time series, denoted by $k_l \in K$, such that $\beta_{k_l} \neq 0$ and $\beta_j = 0, \forall j > k_l$. To choose the optimal lag order in the VAR model Schwarz (1978) suggested the following information criterion (SBC):

$$SBC = \ln(det\,\hat{\Omega}_j) + j\frac{n^2 \ln T}{T}, \qquad j = 0, \ldots, K$$

where $\hat{\Omega}_j$ is the maximum likelihood estimate of the variance-covariance matrix Ω when the lag order used in estimation is j. T is the sample size. The aim is to estimate k_l by the j that minimizes the above criterion. An alternative information criterion introduced by Hannan and Quinn (1979) (HQC)[5] is:

$$HQC = \ln(det\,\hat{\Omega}_j) + j\frac{2n^2 \ln(\ln T)}{T}, \qquad j = 0, \ldots, K$$

Hacker and Hatemi-J (2001) showed that these two criteria perform well in choosing the optimal lag order. However, there are situations when SBC has the highest rate of picking the right lag order compared to HCQ and there are also situations when HCQ outperforms SBC. Nevertheless, unlike Monte Carlo simulation studies, the true model is not known in empirical research. Thus, if these two information criteria pick two different lag orders it is difficult to know which criterion one should rely on. One suggests combining these two criteria to obtain the following information criterion:

$$HJC = \ln(det\,\hat{\Omega}_j) + j\left(\frac{n^2 \ln T + 2n^2 \ln(\ln T)}{2T}\right), \qquad j = 0, \ldots, K \quad (2)$$

A Monte Carlo simulation experiment will be performed to investigate the performance of the above information criterion for choosing the optimal lag order in stable and unstable VAR models. The design of this simulation procedure is presented in the next section.

III. THE DESIGN OF THE MONTE CARLO SIMULATION

The simulations are conducted using the following bivariate VAR(2) model:

Table 1. *Parameter values for VAR model of Equation 3*

$\beta_{1,11}$	-1	0.5	0	0.5	1
$\beta_{1,22}$	-0.5	-0.25	0	0.25	0.5
$\beta_{1,12} = \beta_{1,21}$	-0.5	-0.1	0	0.1	0.5
$\beta_{2,11}$	-0.8	-0.2	0	0.2	0.8
$\beta_{2,22}$	-0.6	-0.1		0.1	0.6
$\beta_{2,12} = \beta_{2,21}$	-0.5	-0.1	0	0.1	0.5

$$\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} 1.0 \\ 1.0 \end{bmatrix} + \begin{bmatrix} \beta_{1,11} & \beta_{1,12} \\ \beta_{1,21} & \beta_{1,22} \end{bmatrix} \begin{bmatrix} X_{1t-1} \\ X_{2t-1} \end{bmatrix} + \begin{bmatrix} \beta_{2,11} & \beta_{2,12} \\ \beta_{2,21} & \beta_{2,22} \end{bmatrix} \begin{bmatrix} X_{1t-2} \\ X_{2t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \quad (3)$$

where the error terms vector $\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$ is normally distributed with ε_{1t} drawn independently from ε_{2t}. To make the results more representative we take into consideration 12 500 ($5 \times 5 \times 5 \times 5 \times 4 \times 5$) possible combinations for the coefficient matrices shown in Table 1. Since $\beta_{2,22}$ is never zero, the VAR model is always of the second order. For each combination set of the parameters 1000 iteration were run. In total the number of simulations equal to 12 500 000 (i.e. $12\,500 \times 1000$). The results are presented on average.

Naturally some of these combinations result in stable VAR models and others result in unstable ones. Whether the VAR model is stable or not it has important implications. The standard asymptotical distributions are only valid for stable or stationary cases. For unstable or non-stationary cases usually the standard distributions are not good approximations and some modifications have to be made. However, since many economical and financial time series are non-stationary it is important to investigate unstable cases also. To separate stable cases from unstable cases we calculate the modulus of the following companion matrix:

$$C = \begin{bmatrix} \beta_{1,11} & \beta_{1,12} & \beta_{2,11} & \beta_{2,12} \\ \beta_{1,21} & \beta_{1,22} & \beta_{2,21} & \beta_{2,22} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

If the modulus of each eigenvalue of C is less than one, then the system is stable.

IV. THE RESULTS OF THE SIMULATIONS

Table 2 presents the frequency distributions for the information criteria presented by equation (2). This results show that the rate of choosing the right lag order is 85.5% for

[5] It is well known in the literature that both SBC and HQC are consistent. By consistency is meant that the criterion selects the true order of the VAR with probability one asymptotically. Nielsen (2001) shows that the consistency results for each criteria holds regardless of the assumption about the characteristic roots in the VAR model.

Table 2. *Results for stable and unstable cases based on simulations for a VAR(2) model*

	Lag length							
Information criterion	0	1	2	3	4	5	6	7
Frequency distribution of estimated VAR orders, $T = 40$								
HJC, stable VAR	4.25	5.0	85.5	5.5	0.9	0.3	0.2	0.2
HJC, unstable VAR	0.0	4.4	87.2	4.9	1.1	0.4	0.2	0.2

HJC signifies the Hatemi-J information criterion presented by Equation 2.

stable VAR models and 87.5% for unstable VAR models. Forty observations were used in the simulations because in many applied studies the number of observations that is available is rather few. However, simulations were also conducted using 100 observations. The results, not presented but available on request, showed that the percentage point of choosing the right lag order even increased (over 90%) regardless of stability or instability assumption of the VAR.

The mean lag was also calculated, which was 1.95 for stable VAR models and 2.03 for unstable VAR models.

V. CONCLUSION

Since the VAR model is intensively used in applied research more research on its properties is warranted. One crucial aspect of the VAR model is the choice of the optimal lag order because all inference in the VAR model is based on the chosen lag order. In the literature there are a number of information criteria that are available for this purpose. However, sometimes these criteria choose different lag orders and then the question is which one should be chosen. Two of the most successful criteria according to the simulation results presented in the literature are Schwarz (1978) Bayesian information criterion and the Hannan and Quinn (1979) information criterion. However, the previous studies show that each of these two different criteria can perform better than the other depending on the properties of the true VAR model. This study suggests combining these two criteria in order to achieve maximal probability to choose the optimal lag order. Another advantage of this procedure is that there will be only one criterion to be used.

The conducted simulation experiments show that this new information criteria can pick the true lag order in more than 85% of cases for small sample sizes ($T = 40$), no matter if the VAR model is stable or not. For larger sample sizes the percentage of picking optimal lag order seem to increase.

REFERENCES

Hannan, E. J. and Quinn, B. G. (1979) The determination of the order of an autoregressive, *Journal of the Royal Statistical Society*, **B41**, 190–5.

Hacker S. and Hatemi-J, A. (2001) Optimal lag length choice in the stable and unstable VAR models under situations of homoscedasticity and heteroscedasticity, unpublished manuscript.

Lütkepohl, H. (1985) Comparison of criteria for estimating the order of a vector autoregressive process, *Journal of Time Series Analysis*, **6**, 35–52.

Lütkepohl, H. (1991) *Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.

Nielsen, B. (2001) *Weak Consistency of Criterions of Order Determination in a General Vector Autoregression*, Oxford University Press, Oxford.

Schwarz, G. (1978) Estimating the dimension of a model, *Annals of Statistics*, **6**, 461–4.

Sims, C. (1980) Macroeconomics and reality?, *Econometrica*, **48**, 1–18.